DOCUMENT RESUME

ED 069 675 TM 002 126

AUTHOR	Adkins, Dorothy C.
TITLE	Objective Measurement of Emerging Affective Traits in Preschool Children.
INSTITUTION	Hawaii Univ., Honolulu. Coll. of Education.
SPONS AGENCY	Office of Economic Opportunity, Washington, D.C. Community Action Program.
PUB DATE	6 Sep 68
NOTE	35p.; Paper presented at American Psychological Association annual meeting (80th, Honolulu, Hawaii, Sept. 6, 1972)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Affective Tests; Attitude Tests; Cultural Differences; *Evaluation Techniques; Factor Analysis; Measurement Instruments; Measurement Techniques; *Motivation; Personality Tests; *Preschool Children; Response Style (Tests); Speeches; Statistical Analysis; Tests; Test Validity
IDENTIFIERS	Gumpgookies; Headstart

ABSTRACT

An objective measure of motivation to achieve for preschool children called Gumpgookies is described. It is an objective-projective technique that requires choice between two alternate types of behavior portrayed in pictures and accompanying verbal descriptions. Gumpgookies are amoeba-like creatures who behave in ways intended to show differences in motivation. The history of the development of the test is reviewed and results of extensive testing reported. Factors analyzed were school enjoyment, self evaluation, purposive behavior, self confidence and instrumental activity. Ethnic-cultural differences are examined. A major problem with the development was in regard to response sets which led to the conclusion that these may be more important in other tests than is recognized. The paper concludes with a brief description of initial development of using the Gumpgookie technique to measure other traits in the affective domain such as warranted self-esteem, warranted other esteem, and integrity or responsibility. (DJ)

ED 069675

EDUCATION RESEARCH AND DEVELOPMENT CENTER

COLLEGE OF EDUCATION
UNIVERSITY OF HAWAII
HONOLULU, HAWAII

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Center for Research in Early Childhood Education
Dorothy C. Adkins, Director

Objective Measurement of Emerging Affective Traits in Preschool Children

Dorothy C. Adkins, Professor and Researcher

TM 003 126

Some of the research reported herein was performed pursuart to Grant Number CG 9929, as well as to Contract Number OEO 4121, with the United States Office of Economic Opportunity. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment on the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official position or policy of the Office of Economic Opportunity.

A paper presented at the meeting of the American Psychological Association, Honolulu, Hawaii, on September 6, 1972.

Objective Measurement of Emerging Affective Traits in Preschool Children*

Dorothy C. Adkins

University of Hawaii

The topic of this address, conjured up in a Chicago snow storm, stresses objectivity of measurement, because lack of it has been a prominent weakness in the affective domain and I wanted to avoid treatment of devices that call for time-consuming content analysis. Stress on objectivity, however, is not indicative of unconcern with other aspects of reliability. By affective traits are meant internally consistent qualities of personality and character dominated by interests, attitudes, appreciations, values, emotions--complex qualities that broadly can be subsumed under such terms as motivation and even morality. The word "emerging" reflects the evanescent quality of affective traits, especially in young children, such traits being subject to development and modification through learning. Finally, I chose to concentrate on young children, for whom need for assessment measures is paramount, but not because affective measurement problems have been solved for older children or adults.

---

*Invited address presented at the meeting of the American Psychological Association, September, 1972. Parts of many of the studies referred to herein were supported by grants to the University of Hawaii from the United States Office of Economic Opportunity, and reports on such studies have already been submitted to that agency. Points of view or opinions stated do not necessarily represent official position or policy of the Office of Economic Opportunity.

You will have sensed already that this is to be not a prêcis but a disquisition. If a prêcis is what you seek, you should not attend invited lectures.

Before proceeding, I should recognize the contributions of my former colleague and continuing close collaborator, Bonnie L. Ballif, now of Fordham University. I am also signally indebted to many staff members of the now moribund Center for Research in Early Childhood Education of the University of Hawaii--Renato Espinosa, J. Michael O'Malley, Frank D. Payne, Phyllis Loveless, June Kimura, to mention only a few--and to David G. Ryans, Director of the University's Education Research and Development Center. Of necessity, I have drawn heavily upon earlier reports and articles to which one or more of the above-named persons have contributed.

You should be warned that you will learn more about Gumpgookies, a test of young children's motivation to achieve in school, than you may care to know. If you have heard of this test before, some redundancy will be necessary. But this audience should not need to be reminded that repetition is the second law of learning. (Maybe E. L. Thorndike did not say that, but he might have!) Besides, if Allen Edwards were talking you could expect to hear about social desirability, not little gumpgookies. So you have made a choice of which you may have been unaware.

Literature relevant to the topic will now be reviewed. In brief, there is none. Although thousands of references could be cited on objectivity, measurement, affective traits, and young children, no treatments are germane to the entire topic. Hence without further ado about nothing, we turn to exegesis of the topic.

Motivation to achieve in school has been conceptualized as a hypothetical construct that explains aspects of achievement-oriented behavior not attributable to intellectual abilities. It appears to be determined by a combination of attitudes, feelings, or expectations--covert responses that can be learned. Five types of covert responses have been hypothesized as essential components of motivation to achieve (Adkins & Ballif, 1970b, c).

The first is expecting affective or hedonic change. The young child must expect that if he engages in achievement-oriented activity within the school setting his life will be more pleasant.

The second constituent focuses on the concept of self as an achiever in learning. Perceptions about the self appear to be crucial in the causation of behavior, the feeling of personal adequacy being of pervasive importance in the child's perceptual organization and functioning in the classroom.

A third component arises from the direction or purposiveness of behavior implied in the concept of motivation itself. It is, in essence, the setting up of purposes for the self-direction of behavior. These goals often go beyond the immediate moment and suggest implications for future times and situations.

Closely related to purposiveness of behavior is knowledge of instrumental steps that will be effective in accomplishing purposes. The first instrumental step toward any purpose is realization of personal responsibility for action and of personal control over outcomes. An individual must believe that some action on his part helps or is required to result in the desired goal. In addition,

he must know that he should autonomously initiate work activity instrumental in accomplishing his purposes.

The last hypothesized component is self-evaluation. In addition to a positive self-concept, self-assessment or self-evaluation is essential. This process requires not only presence of an internal standard of excellence but also comparison of actual performance with this standard.

Achievement-motivated behavior, then, is regarded as a result of dynamic interaction of learned responses. Motivation to achieve in school will be evident only when a child expects that achieving in school will be pleasant; thinks that he can achieve in school; can set up his own purposes to achieve; knows the instrumental steps that will lead to his achievement; and can evaluate his own performance against an internalized standard of excellence. A summary of the literature documenting that the types of responses considered here are subject to learning and therefore may be taught has been presented elsewhere (Adkins & Ballif, 1970c). It will not be repeated here, where principal concern is with how to measure components of motivation and, more broadly, other traits in the affective domain.

The task on which I embarked in 1965 and on which I was joined by Ballif in 1966 was to develop a testing procedure that would not only accurately measure evasive components of motivation to achieve but also be effective within the limited response repertoires of preschool children. Probably the most influential approach to measuring achievement motivation has been the work of McClelland,

Atkinson, and their associates (McClelland et al., 1953), who have used fantasy as the medium through which themes, needs, and goals are scored for achievement content. Despite the appeal of this idea and the fact that a reasonable degree of objectivity or rater reliability can be achieved, the generalizability of scores across content and across time is open to serious reservations, even for adults, as Entwisle has recently documented (Entwisle, 1972). Further complications arise when such a procedure is attempted with very young children. Many tend to withdraw in the testing situation, and the majority lack verbal skills needed to describe fantasies. Moreover, absence of universal child-rearing practices means that young children have not been exposed to uniform experiences, so that both their understanding of picture stimuli and the content of their fantasies are limited.

Extensive search for an appropriate method of measurement included a variety of techniques and formats covered in previous reports (Adkins & Ballif, 1968, 1970c). From these initial endeavors, sufficient direction was obtained for a new measure of motivation to achieve, Gumpgookies. It is an objective-projective technique that requires choice between two types of alternative behavior portrayed in pictures and accompanying verbal descriptions. It centers around activities of imaginary little figures called gumpgookies. The gumpgookies behave in ways intended to show differences in motivation to achieve in activities appropriate for young children. Each item presents two gumpgookies in a semi-structured situation. The child is told that he has his own gumpgookie and

that, although it looks like all others, it follows the child and behaves as he behaves--it likes what he likes and does what he does. As the examiner points to each illustration as it is described, the child selects and points to his own gumpgookie. For example:

This gumpgookie does what it wants to.

This gumpgookie does things well.

Which is your gumpgookie?

A gumpgookie is an amoeba-like character that, although faceless, has a suggestion of a head, two arms, and two legs. (Just as James Stewart believed in Harvey, so do Ballif and I and our hundreds of young children believe in gumpgookies.)

From some 300 items, 200 items were selected for the first form. For each item, the two gumpgookies appeared side by side, the left one being described first. This instrument was administered in two sittings to 182 preschool children. Approximately 90 of the children were selected by pooling judgments of a teacher and two aides as to the child in their class most motivated to achieve and the child least motivated to achieve.

A measure of the relation of each item with the total score and a discrimination index for the external criterion (i.e., high versus low motivation) were obtained. The matrix of inter-item phi correlation coefficients was factored by the principal-axes method and the factor matrix rotated to oblique simple structure by a biquartimin solution with gamma equal to .5. The eigenvalues had not decreased to unity even when as many as 20 factors had been extracted. Since there was no hope that so many factors could be

interpreted, the number was set arbitrarily at six or seven and, for some solutions, at three.

Rotation of the initial set of seven factors permitted only extremely tentative identifications, which in turn provided only limited evidence for the hypothesized constituents of motivation to achieve. Inter-correlations among the seven factors were also factored, yielding a three-factor, second-order matrix that was also rotated to oblique simple structure. This analysis provided a somewhat clearer three-factor structure. In view of subsequent developments, however, interpretation of these factors will not be presented.

At this point, Gumpgookies was revised to consist of 100 items and was administered in one sitting to a new sample of 330 children. Data again were analyzed in terms of basic test statistics; and, although factor-analysis techniques were applied, a number of alternative approaches also were pursued. One was designed to yield clusters with maximum K-R 20 reliability estimates, for which Joseph Klock provided a program. Results of this analysis were rather similar to those of factor-analytic methods. Moreover, anomalous results, such as negative reliability coefficients, sometimes occurred, and a modification of the program that was possibly needed was not then available. Hence this technique was abandoned.

Another approach, Congor's dimensional analysis of binary data, was brought to our attention by Ledyard Tucker. Consideration of this method, however, led to the conclusion that it would lead to about the same results as more traditional factor-analytic techniques.

The prospect of difficulty factors in analyses of binary data was not unknown. After discussions with Paul Horst, however, the decision was to proceed with factor-analytic techniques and attempt to interpret factors that hopefully would transcend the influence of difficulty.

Although the answer key for the original 200-item form had been determined in a random order, half left and half right, in the original key for the 100 selected items an unusually large number had answers corresponding to the right-hand illustrations, which also coincided with verbal descriptions read last by the examiner. This discovery, however startling, was not inconsistent with the fact that improbable events do indeed occur, with predictable relative frequencies. Suspicions had been aroused, but vacation periods and demands for a revised form of the test were imminent. Accordingly, the key for the 100 items in the revised form was again randomized between left and right. (This early history will be familiar to some of you, but I can scarcely assume that all of you have read everything we have written.)

Further study of factor and cluster analyses of data on the 200-item form and on the first 100-item form soon revealed curious problems. Certain factors or clusters had most keyed answers in the right-hand position, others in the left-hand position. With the test format used, the left-right and primacy-recency influences were inextricably confounded, as noted above.

Three principal approaches were pursued in efforts to understand this problem (Adkins & Ballif, 1970a). One was to divide

answer sheets into two groups--one that did and one that did not differ significantly from the number of runs (successive responses of right or left answers) appearing in the answer key. Data for the two groups of subjects were then separately factor-analyzed. Without presentation of agonizing details of the analyses, it must be reported that outcomes were inconclusive. The most probable explanation was that the statistical criterion used to separate subjects into those susceptible and those not susceptible to runs was not well adapted to detections of subtle psychological influences that determine what on the surface appear to be erratic shifts of set among preschoolers, given the original format and nature of this particular test.

A second attack on the problem of set factors yielded more definitive results. Artificial score matrices, with randomly-assigned equal numbers of answers in each position, were constructed for 24, 30, or 36 subjects. The answer patterns and item inter-relationships were designed so as to yield two factors, some very clear ones and others weak. Then the original answer patterns were overlaid with complete position preferences (or the equivalent primacy or recency preferences) for varying numbers of subjects. With strongly determined factors, imposition of position preferences for roughly a fourth of the subjects altered the original factor structure to position factors only; i.e., resulting factors had answers appearing in only a single position. With weaker initial factor structure, overlaying position preferences on the answers of an even smaller

fraction of subjects (perhaps a fifth or a sixth) shifted the factors to dominance by answer position.

Even though the straightforward nature of shifts in answer patterns in matrices analyzed by the foregoing means differs from less easily discernible patterns characteristic of responses of the four-year-old children on whom the original work had been done, this second approach confirmed that position factors had to be contended with.

A third method that confirmed position preferences was rearrangement of the inter-item matrix of phi coefficients so that sets of items with correct answers at the right and at the left each appeared together. Almost without exception, mean coefficients of items with others having the same answer position were positive; with others having the reverse answer position, negative. Mean positive coefficients were almost uniformly larger than mean negative ones, however.

The finding of more than one right factor and more than one left factor indicated that some content variables were involved. This optimism was bolstered by the fact that many of the items did discriminate between children selected by teachers as having high and low motivation. Moreover, mean total scores of four-year-old Head Start children in a group composed of those selected by the teachers as the three most highly motivated and those selected as the three least highly motivated differed significantly. Further confirmation lay in the fact that score distributions, even for the youngest groups, did not fall equally below and above a score

equivalent to 50% of the items but started at or near the 50% score and progressed upwards. And, in general, mean scores increased with increasing age. The first calculation of the correlation of Gumpgookies scores with IQ yielded a significant r of .31, which again was interpreted to mean that factors other than chance were operating.

For testing of several ethnic-cultural groups scheduled for 1969, the 100-item test was revised further: (a) positions of the illustrations were no longer confined to left and right but also included up and down, lower left and upper right, and upper left and lower right; (b) order of description of figures was randomly determined; (c) answer positions again were randomized, taking into account both position of the illustration for the keyed answer and order of presentation; (d) wording of many items was simplified to reduce cognitive and verbal difficulty; (e) items objectionable for one reason or another were removed; and (f) the test was shortened to 75 items. Two main forms of the test resulted, one for individual administration to preschool children and one for group administration to first- and second-graders. These are the forms from which data reported later were derived.

In retrospect, efforts to get rid of effects of response sets simply by means of revising the format were not successful. Extraneous influences had only become somewhat more difficult to detect. Parenthetically, these response sets have no systematic undesirable influence on total test score, because the subject is expected to get only a chance score on items to which he responds on the basis

of a particular set. But response sets do affect items loaded on particular factors, so that a subject could get unwarrantedly high or low scores on separate factors. Moreover, effects of response sets on the composition of the factors made interpretations tenuous.

Since change in format had not been successful, another solution to the response-set problem had to be found before factors could be interpreted with any assurance. The next approach was to obtain response set scores for each subject, partial these out of the item intercorrelation matrix, and then factor (Adkins & Ballif, 1972). For each subject were computed the numbers of his answers that were in the left-hand position, that were in the up position, and that had been presented first. For items in which alternatives had been placed in a diagonal position, e.g., upper left and lower right, an arbitrary decision was made to regard upper left and upper right as up, lower left and lower right as down. This was done because the small numbers of items with answers in the two diagonal positions would have resulted in response-set scores of very low reliability for these positions.

The mathematical solution for partialling out these three variables was developed by Horst, and the computer program to effect the solution was worked out by Renato Espinosa and Robert Bloedon, members of the Hawaii Center staff, with Horst's guidance. It yields orthogonal factors completely uncorrelated with response-set scores (Horst, 1972).

The complete program provides, among other things, correlations of each item with response-set scores; rotated "partial" factor

loadings for each item; and reliability estimates (K-R 20) for total score, partial factor scores, and response-set scores. It also yields exact factor scores for each subject, based upon regression weights for each item.

Separate analyses have been made for 1813 four-year-old children, for 10 separate ethnic-cultural subgroups of four-year-olds, for 128 first-graders, for 122 second-graders, for 250 first- and second-graders combined, and for a total group of 2313 children. The K-R 20 values for the partial factors tend to be higher for the older children, and those for partial factors tend to be less chan for factors based on the zero-order correlation matrix. This is doubtless true because the latter factors include reliable effects of response sets. Response-set scores are more consistent for the older children. Influence of a primacy-recency set is relatively greater for older children, while younger children are more prone to answer-position sets.

Details of extensive work on comparing several solutions for different numbers of factors and for different groups, as well as in comparing partial factors and unpartial factors, will not be presented here (Adkins & Ballif, 1972). It soon became apparent, with respect to both the original unpartial factors and the partial factors, that those for the four-year-olds do not correspond to those for the first- and second-graders very closely. It was not unreasonable to suppose, however, that the factorial composition of motivation to achieve in school changes with age. Indeed, such is almost certainly the case. Yet, despite the conviction that

changes with age in the factors affecting the test responses were to be expected, attempts to interpret the changes have not been pursued at length because of the small amount of data for older children.

Full exploration of this problem led to question as to the dependability of factor loadings obtained from phi coefficients based upon relatively small numbers of cases. Although the original plan was to have at least 200 cases for any factor analysis, probably this number was too small. Hence certain samples were divided at random into halves and separate factor analyses were made for each half as well as for the total sample. The similarity of the three sets of factor loadings for each sample was investigated by inspecting the correlations of the loadings from the three solutions, i.e., for the two half samples and for the total sample. A factor for the total sample was regarded as verified when a factor in one half sample and a factor in the other half sample each showed its highest correlation for the same factor in the total sample while these same factors for the half samples had the highest correlation of any pair of factors across the half samples.

Detailed results of applications of this approach are presented in a forthcoming article (Adkins & Ballif, 1972). Somewhat later, at the urging of Tucker and Harry Harman, congruency coefficients instead of correlation coefficients were compared, with substantially the same results.

Perhaps the most defensible interpretation of factors results from the five-factor analysis based upon 2313 cases, including

2063 four-year-olds and 250 first- and second-graders. Although the K-R 20 estimates of reliability for the total test score on Gumpgookies have been in the neighborhood of .83 to .93, the estimated coefficients for the five factors are not so high, ranging from .35 to .55 for the large combined sample. This is not surprising, since the total test consists of only 75 items.

For the interpretation of a factor, the method has been first to list the items that have their highest loading on it for the total sample. Then the loading of that item for the corresponding factor in each half sample is recorded, with a notation as to whether it is the highest loading for the item. Greatest weight is accorded those items verified in all three analyses, i.e., items for which the highest loadings apply to the appropriate verified factors. Attention is also given to size of loadings.

Factor A consists of items indicating an autonomous activity orientation permeating use of time and interaction with others. This "on-the-go" behavior is more than generalized activity; it is initiating and engaging in specific behavior that is appropriate to insure success in the particular tasks and situations at hand. It involves knowing the effective instrumental steps and taking them. These activities are instrumental to achievement in general, e.g., wanting to work longer; to achievement in school, e.g., keeps trying to write numbers; as well as to obtaining reinforcement for achievement, e.g., shows its paintings to others. The factor is referred to as Instrumental Activity.

The reflection of a preference for school- and teacher-related experiences is clear in factor B, a School Enjoyment factor. Specific items include wanting to go to school to learn and liking learning, along with watching and helping the teacher as opposed to engaging in other activities. This positive attitude toward school is further exhibited by identification with the teacher, e.g., wanting to be the teacher when playing school.

The items constituting factor C, a Self-Evaluation factor, represent ability to evaluate one's own performance coupled with confidence that the evaluation will be high. The process of self-evaluation is suggested by items portraying gumpgookies who know when their work is right, when they are doing well in school, what they can and cannot do, and whether or not they are always doing their best. Items describing gumpgookies who are self-evaluated as always at their best and doing well also suggest a feeling of their own excellence.

Factor D consists almost entirely of items set in competitive physical situations, e.g., winning in running, climbing higher, and leading in follow the leader. Apparently it represents Self-Confidence in coming out on top, in being the best or better than others. With additional items staged in other settings, the factor probably would transcend physical activities. Indeed, for another analysis based on the 1813 four-year-olds, emphasis of physical activities in a factor interpreted as self-confidence was reduced.

The common denominator for items loading on factor E, a Purposive Behavior factor, has to do with awareness of future

implications of present behavior. The gumpgookies in these items are still trying to obtain future goals, e.g., trying to write, apparently being directed by self-initiated purposes.

Ten subgroups within the 1813 four-year-olds could be identified. They are referred to loosely as ethnic-cultural groups, comprising Mormons, Catholics, Jews, American Indians, Mexican-Americans, orientals living on the west coast of the United States, residents of Hawaii (not by any means all pure Hawaiians), urban blacks, rural whites, and Puerto Ricans. I can be the first to find fault with our sampling. The majority of the children were enrolled in Head Start classes and came from homes of low socioeconomic status. It was not possible, however, to locate conveniently groups of Mormon, Catholic, and Jewish children from homes of low socio-economic status. Significant portions of certain samples had been exposed to a language other than standard English. There was no systematic control or variation of the rural-urban dimension. Nonetheless, results both with respect to substantive factors and response sets may be suggestive (Adkins, Payne, & Ballif, 1972).

For the age range in question, a small positive correlation with age was found again for total score (.34) and somewhat lower correlations for all five exact factor scores. (Observe, parenthetically, that a zero relation with age for a test of motivation would be very suspect, while a high relation might well mean that the test is measuring general mental ability.) Although the correlations were small, their effects were removed in a procedure that yielded

age-normed Z scores (Adkins & Payne, 1971). Five 2 x 10 (sex by ethnic-cultural group) analyses of variance were performed using the fixed effects model.

The 10 groups differed substantially in total score. Boys and girls did not differ significantly, although such a difference was not precluded in the test development procedures, as it is for the Stanford-Binet. The three middle-class samples--Mormons, Catholics, and Jews--had higher total scores than the lower-class samples. Mexican-American, West Coast Oriental, American Indian, and Hawaii samples had the lowest mean scores. On Instrumental Activity, although the middle-class samples had relatively higher scores than the majority of lower-class samples, the Puerto Rican sample was second only to the Catholic group. American Indian, Hawaii, and Mexican-American samples again had the lowest mean scores.

A significant but weak tendency emerged for girls to exhibit higher scores on School Enjoyment than boys. This tendency held for all groups except the White-Rural and Oriental (West Coast) samples, which contained few subjects. These results support the conclusion that girls at this age, regardless of ethnic-cultural membership, enjoy school slightly more than do boys.

The ethnic-cultural groups also differed significantly on School Enjoyment, although the percentage of variance accounted for was not large. The relative standings of the groups run contrary to any categorization on the basis of socioeconomic status, urban-rural dichotomy, or geographic region.

For boys and girls combined, the Negro-Urban sample exhibited higher mean School Enjoyment scores than either Jewish or Catholic samples. In fact, the mean score for the Catholic group ranked only sixth among the 10 groups. For girls, Mormon and Jewish samples exhibited the highest mean scores; among boys, the Mexican-American, Catholic, and Puerto Rican samples had lowest mean scores.

For Self-Evaluation, only ethnic-cultural membership produced significant differences and the percentage of variance accounted for was higher than that for either of the first two factors. The three middle-class samples had the highest mean ability to evaluate their own performance, while the Mexican-American, Oriental (West Coast), American Indian, and Hawaii groups had the lowest mean scores.

Significant ethnic-cultural differences emerged on the Purposive Behavior factor. Although significant sex differences were not obtained, there was a slight tendency for boys to score higher than girls. The highest mean scores were obtained by Jewish boys and girls and by White-Rural and Oriental (West Coast) boys. The Mexican-American and Negro-Urban children, as well as the Oriental (West Coast) girls, obtained the lowest scores.

Early on, before a method of reducing effects of response sets on factor composition had been developed and before there was full appreciation of the need for very large samples of young children to determine factors in the affective domain, we had done separate factor analyses for each of the 10 ethnic-cultural groups. Many hours were devoted to attempts at interpretation; to comparisons

of factors among groups; and to study of differences in nature, extent, and effects of response sets among the groups. Results of these efforts were discouragingly inconclusive. More recently Myrna C. Ibarra, a graduate student, has applied the method of factoring with response-set scores partialled out to the seven largest groups, realizing that N's in the neighborhood of 200 were still undesirably small. She obtained congruency coefficients among 35 orthogonally rotated "partial" factors, five for each group. To interpret this matrix, she factored it for varying numbers of factors from five through 10. The seven-factor solution appeared best, so the initial factors which had the highest loadings on each of the seven factors were examined, the items highly loaded for each being listed. Strong verification across all seven groups was found for an Instrumental Activity factor and a Purposive Behavior factor. Results of this type of approach are still inconclusive as far as the other posited factors are concerned--perhaps because of small N's but also possibly because factor structure does indeed differ among the groups.

Let us return more specifically to evidence regarding response sets. For a five-factor solution based on the 1813 preschoolers, the Up-Down score correlated .7[illegible] with loadings on one original, i.e., "unpartial" factor, -.34 for another, and -.41 for a third. This means that items on the first factor were predominantly up-keyed items, those on the other two tending to be down-keyed items. The Left-Right score correlated -.79 with loadings on one factor. For the first- and second-graders, the Primacy-Recency score

correlated .64 with loadings on one factor, -.30 with those on another. For this group, the highest correlation for either of the position scores was -.29 for the left-right score.

Comparison of means and variances of response-set scores among the eight largest of the ethnic-cultural groups is of some interest. Although there were some significant differences between pairs of means, especially in a tendency for American Indian and Hawaii children to slightly prefer the down to the up position, in contrast with the other groups, the differences are small.

More striking are differences in variances, those groups with higher mean scores on the total test being significantly less variable on response-set scores. This finding is not surprising, because the groups differed in mean scores on the total test and on the factors. Individuals who on the whole find the items difficult are likely to respond in accordance with response sets. Hence standard deviations of set scores for higher-scoring groups tend to be lower than those for lower-scoring groups.

While on the average no prominent response sets favor either primacy or recency or certain answer positions, some children are affected by particular response sets, some making responses they hear first, some those they hear last, and some those in each of the answer positions in question.

The K-R 20 reliability estimates were examined separately for the eight largest ethnic-cultural groups. Especially striking are the relatively high values for left-right and up-down scores for the four groups that were lowest on the total test (Mexican-American,

Hawaii, American Indian, and Puerto Rican). The reliability estimates thus are consistent with data on standard deviations.

Pleased though I have been with the notion of partialling out response-set scores to yield factors uncorrelated with them, misgivings assail me every now and then, especially when I realize that a subject with a very high response-set score may get a very high score on one or more of the factors. This effect, to be expected by the very nature of the technique, is strikingly revealed by plots of response-set scores against exact factor scores. I am now exploring application of a linear correction, whereby a constant times the sum of the absolute values of the three response-set scores is subtracted from the exact partial factor score. This procedure yields corrected scores that have negative correlations with the response-set scores, which are intuitively appealing.

Another approach I have only recently used is to reduce the total sample to about three-eighths of its original size by discarding subjects whose sum of absolute values of response-set scores exceeds some small arbitrary value. The data for the surviving sample are then factored by the ordinary method. The hypothesis is that the resulting factor structure will closely resemble that obtained for the full sample by the "partial" factor method. For whatever reasons, the resemblance does not appear to be as close as was expected--perhaps because the structure for the reduced number of cases is too unstable, as revealed by attempts to verify it across half samples. (I may now confess, also, that I am always skeptical of results that are

spewed forth from the giant machines. I had complete confidence in the accuracy of the 22,500 phi coefficients that appeared as a footnote to my dissertation, but I lack this feeling of certitude with respect to outputs of high-speed computers--and fairly often with good reason.)

Another recurring idea with considerable appeal is not directly applicable to Gumpgookies items in their present format, which involves different illustrations for the two options for some items. With identical illustrations and wording of options so that each is independently meaningful, however, one could assemble, say, eight forms of the test. Each item would appear in eight guises, the keyed answer appearing twice in each of four positions--up, down, left, right--and in each position one time being presented first and one time last. Each form would be given to some 200 or 300 subjects, results amalgamated, and the matrix factored. Factors so obtained should be free of effects of response sets. Implementation of this idea must await the largess of one of the great federal spenders.

The attention given to response sets here is warranted by the likelihood that young children's performance on many other instruments must also be affected by similar processes. Persons developing tests for children in both cognitive and affective realms cannot sensibly ignore this problem.

Do the response sets have significant meanings in their own rights, as has been argued for such sets as acquiescence and social desirability? Quite possibly. Perhaps a tendency to choose the first

answer presented suggests impulsivity, low auditory attention span, lack of patience. The recency tendency may indicate patience or restraint, a longer auditory attention span, or even curiosity. Recall that neither trait is dominant for either the preschoolers or the first- and second-graders, but that for Gumpgookies the influence in one or the other direction is distinctly greater for the older children.

Does a tendency to take the alternative presented at the left reflect some sort of vicarious reading habit--vicarious because most of the four-year-olds do not read but may have been read to--, perhaps a short visual attention span if the choice looked at first is in the left-hand position? Is a tendency to choose the right-hand option influenced by a longer visual attention span, or by the fact that the examiner is at the right of the child or usually is recording with her right hand?

Similarly, is an up choice affected by vicarious reading habits, a short visual attention span, or even possibly by optimism? Does a down choice reflect pessimism, laziness in that the down figure is easier to point to? We do not know.

We do know that, for each of the three sets by which children may respond when a choice is too difficult, there is no universally dominant tendency in either direction. But, especially considering the small numbers of items involved, the tendencies are reliable for subjects who find the test difficult. For example, the up-down scores show a K-R 20 of .79 and the left-right scores a K-R 20 of .64 for our Mexican-American sample.

For another haunting problem we have not been able even to attempt a solution. Certain factors, both partial and original unpartial ones, exhibit strange relations with item positions. One factor may be loaded with items predominantly in the first third of the test, another with items concentrated toward the end, still another with items scattered throughout. Does the first case represent initial but short-lived enthusiasm, the second increasing interest or possibly learning as the test progresses, the last a dogged persistence or even an end-spurt effect? Moreover, items contiguously situated in the test seem to cluster on factors. Is this effect attributable to their general position in the test, to the fact that they are contiguous, or to chance placement in the test? We know how to find the answers to such questions if supporters of educational research are interested in measurement in the affective realm.

The relation of Gumpgookies' item difficulties or endorsement percentages to age has been examined. Another of my students, Ma. Lourdes S. Villanueva, has studied this question intensively in relation to factor loadings of items. In a way, this can be a treacherous endeavor, since some items loaded highly on certain factors may not show age changes in endorsement percentage because of failure of learning-teaching environments. However, when an item shows no age change and in addition has only weak factor loadings, it is a candidate for discard. One item, for example, which required choice between liking one's own house versus wanting a prettier one, showed no age change and negligible factor loadings.

Another item, asking "Which is your gumpgookie climbing?", on close inspection has ambiguous illustrations, since the one higher in the tree appears to be resting. Little age change occurred for this item. Again, choice between "likes to tell stories" versus "likes to listen" shows little age trend coupled with weak factor loadings. We have examined the results, item by item, though I will not keep you for several more hours to present details.

Heretofore I have mentioned tangentially aspects of both reliability and validity. K-R 20 estimates for the total test hover between about .83 and .93, depending on age range. On a few occasions we have been able to compute test-retest coefficients, which have been in the neighborhood of .60 to .70 for both preschool children and first- and second-graders in one-year age ranges.

Content validity is claimed through the construction of items to accord with the general theory. Interpretation of factors affords one type of evidence of construct validity. Low positive correlations with age and Stanford-Binet IQ provide additional information, strengthened by somewhat higher relations with the Caldwell Preschool Inventory, a measure highly correlated with IQ but with greater orientation to achievement.

As for criterion-related validity, which I perhaps old-fashionedly still consider important, recall that for the original 200 items one selection factor involved discrimination indices based on teacher and aide nominations of the most and least highly motivated children in their classes. In several instances test scores have been compared with teacher ratings based on different scales. For the score on

12 selected items from the Zigler Behavior Inventory, administered in full, the rank-difference correlation of .48 was significant with an N of 16. For a scale composed by Ballif, one teacher's ratings correlated .58, a special language teacher's ratings .72. When 10 preschool teachers indicated by rankings the three most and the three least motivated children, 17 of the highest 30 were above the median and three at the median test score. Of the 30 ranked lowest, 10 were above the median and one at the median. Such findings and additional data for first- and second-graders yield differences significant at the .05 level--not to be dismissed lightly in view of the ubiquitous problems with teacher ratings that are especially troublesome when distinctions among aptitude, achievement, and motivation are involved.

Remember, too, that our venture was embarked upon with a conviction that motivation to achieve in school is learned and therefore should be teachable. For several years, the Hawaii Center has worked on development and tryout of special curricula designed to promote motivation, most recently with a small group at Fordham University spearheaded by Ballif. Many problems accompany such endeavors. Teacher N's are small. Some teachers do not understand or apply the designated curriculum. Some adhere solely to sweetness and light, trusting to nature. Some fail to elicit needed cooperation of parents. Certain teachers in comparison classes are more motivating than those in special motivation classes. Other contrasted preschool programs, with particular emphasis on regular daily achievement accompanied by material or social rewards, may be highly motivating,

as we have found with our language and mathematics curricula, for example. The picture is not so bleak as the preceding qualifications may have led you to suspect. But, while our motivation curriculum does indeed produce significant increases in age-normed test scores, so often do some other curricula produce significant increases.

Unfortunately, we have not been able to assign children randomly to treatment versus no treatment--a condition that does not exist in a real world--or even to contrasted treatments. We do not claim that our curriculum enhances motivation more than some other curricula in the hands of some teachers can do. This is no cause for dismay, for it may indicate that a variety of teacher styles and curriculum content can enhance the preschooler's motivation to achieve in school. Nevertheless, we have continued to pursue outcomes of motivation curricular units in terms of both Gumpgookies and other measures more specifically related to the curriculum. Previous reports are available (Adkins & Espinosa, 1971; Adkins & O'Malley, 1971), and our latest findings will be available shortly.

Interlarded with increasing sophistication as to how to cope statistically with data on affective characteristics of young children have been some insights into how to construct items. We now know better how to talk like four-year-olds. (You may have become aware of this!). We avoid contractions. We do not carry context over from one alternative to another. We adhere to the present tense. We use short sentences. We suggest identical illustrations for both choices in dichotomous items. Randomization

of primacy versus recency and of answer position for both total score and factor scores, once the structure has been determined, is indicated. Underlining of key words in both right and wrong answers, to help to control emphasis of examiners, should be the practice. Socially acceptable words ("help," "try," "share") in right answers and socially unacceptable answers (often negatives) in wrong answers are to be used sparingly. Review of all items by persons experienced with young children and preliminary tryout of new items, with intensive queries of subjects, are advisable. Because of uncontrollable influences of different illustrations for alternative answers, some being possibly more appealing than others, my suggestion now is to use identical illustrations for both options in an item and to change the illustrations from one item to another.

Heretofore, to the inspection and interpretation of data on Gumpookies I have devoted what seem to be googols of hours. ("Googol," in case you do not know, is the word for the largest number to which a word is assigned, a one followed by a hundred zeroes.) Yet sometimes I wonder whether or not indiscriminate efforts to increase achievement motivation would be wise. It is reported that a German general Baron von Hammerstein, divided qualities of his officers into four classes--cleverness, stupidity, industriousness, and laziness--, most officers possessing two of these qualities. He felt that the clever and industrious are fitted for high staff appointments and that use can be made of those who

are stupid and lazy. One who is clever and lazy, however, is destined for high command, for he has the temperament and nerve to deal with all situations. "But" (to quote) "whoever is stupid and industrious is a danger and must be removed immediately."

I recall, also, how Marion Richardson and I used to speculate about the need in the federal government for a special agency for tenured employees who were both incompetent and motivated, their only assigned duty being to cash their pay checks. (When advance holding of federal taxes was invented by some larcenous-minded individual, I had one employee who thought this must mean that he should cash only alternate checks. After two years the Department of the Treasury was in a complete swivet.) Hence I have decided to devote some attention to other affective traits that can be subsumed under the broad term "moral development."

Some federal and state government officials have been intimating that the paramount concern of early education should not be cognitive development but character development. Since I agree with this point of view, I propose extension of work on measurement of motivation to cover other traits in the affective realm, chief among which are what I refer to as <u>warranted</u> self-esteem, <u>warranted</u> other-esteem, and integrity or responsibility. To this end, I have constructed a large number of objective-projective test items, 80 of which have been tried out for only roughly a hundred subjects of mean age four. The factor structure even for this first set of items and small number of subjects is highly promising--a clear integrity factor; a factor definitely related to esteem of others or altruism,

including sharing and helping, and a factor related to independence and self-esteem, even to the point of downright lying in order to preserve self-esteem.

The interpretations for this three-factor solution are fairly simple and yet present some problems, especially with respect to certain items on which the predominant tendency seems to be for the child to lie in order to preserve independence or self-confidence. Detailed comparisons of the three-factor solution with four-, five-, and seven-factor solutions were made, with items identified that, where possible with the numbers of factors involved, presented the same patterns. Thereby emerged five factors, which corresponded fairly closely to those in the five-factor solution.

The least ambiguous factor of all can be named "Altruism." It clearly involves sharing and helping behavior--trying to teach others how to play a game, showing a lost one how to get home, getting a bandage for another's hurt toe, sharing lunch, waiting for one's turn, and so on.

Two integrity factors, which merge in a three-factor solution, appear in that for five factors. One entails more social orientation than the other. Choices reflecting the first are, e.g., sometimes playing with others, making presents for others, telling its mother its book is lost, admitting that the teacher wrote its name, admitting that it got dirt on the floor. The second integrity factor seems to imply a sense of personal responsibility for doing one's share or what is regarded as right but with less direct regard for others--either peers or adults. Thus the child high on this factor chooses behavior such as starting to clean up spilled sand, leaving money on a table, returning a toy to its

owner, doing all it can versus wanting others to do work, doing something by itself. This factor reflects a personal standard of honesty and responsibility.

Two other factors that went together in the three-factor solution separated when five factors were extracted. Although both are related to self-concept, they are difficult to distinguish, both involving lying to maintain a positive self-image. One emphasizes work and persistence; somewhat irrespective of reactions of others--finding something to do when sad, not caring if others laugh when it is right, claiming it painted a picture when it did not, claiming it wasn't at fault in breaking a dish. The factor suggests selfishness and dissimulation to preserve a strong self-image. The other of this pair of factors seems to place more reliance on rejection of help but still stresses independence through such choices as liking to build its own house, claiming to build its house itself when in reality it had help, trying until it finishes something hard, claiming to have drawn a picture that was given to it. Note that in both factors the child typically is unable to identify with a character that revealed a fault, such as breaking something, or that did not know something, or that required help.

The data on which these results are based are inadequate to provide firm conclusions for differentiating the factors definitively. But the K-R 20 reliabilities for the exact factor scores range from .65 to .72, higher values than have been found in general for motivation factor scores based upon approximately the same number of items and much larger numbers of cases, while the estimate for the total score is .82.

Although retest reliability estimates are not available, one set of 40 items correlated .69 with another set of 40 items administered some two weeks later.

Some of the new untried items are designed to shed further light on warranted versus unwarranted self-esteem and other-esteem. Myriad opportunities for research on the emergence of constellations of behavior in these important areas of moral development somehow must be created. Once such components are measurable, homes and schools can apply techniques to discourage proliferation of unwanted traits and to enhance development of those desired.

REFERENCES

Adkins, D. C. & Ballif, B. L. Exploration of motivation to achieve in preschool children. Final Report to the Office of Economic Opportunity, 1967.

Adkins, D. C. & Ballif, B. L. Measurement of motivation to achieve in preschool children. Final Report to the Office of Economic Opportunity, 1968.

Adkins, D. C. & Ballif, B. L. Developing a preschool measure of motivation to achieve. In C. E. Lunneborg (Ed.), Current problems and techniques in multivariate psychology: Proceedings of a conference honoring Professor Paul Horst. Seattle: The University of Washington, 1970. (a)

Adkins, D. C. & Ballif, B. L. Factors of motivation in young children: Theoretical and empirical. Educational Perspectives, 1970, 9, 7-11. (b)

Adkins, D. C. & Ballif, B. L. Motivation to achieve in school. Report to the Office of Economic Opportunity, 1970. (c)

Adkins, D. C. & Ballif, B. L. A new approach to response sets in analysis of a test of motivation to achieve. Educational and Psychological Measurement, Autumn, 1972.

Adkins, D. C. & Espinosa, R. Programmatic research on preschool language, quantitative, and motivation curricula combined with parent participation. Report to the Office of Economic Opportunity, 1971.

Adkins, D. C. & O'Malley, J. M. Continuation of programmatic research on curricular modules for early childhood education and parent participation. Final Report to the Office of Economic Opportunity, 1971.

Adkins, D. C. & Payne, F. D. Age-normed "Z" scores for the 75-item randomized individual form of Gumpgookies. Report to the Office of Economic Opportunity, 1971.

Adkins, D. C., Payne, F. D., & Ballif, B. L. Motivation factor scores and response set scores for ten ethnic-cultural groups of preschool children. American Educational Research Journal, Fall, 1972.

Entwisle, D. R. To dispel fantasies about fantasy-based measures of achievement motivation. Psychological Bulletin, 1972, 77, 377-391.

Horst, P. Factor scores independent of item traits. Educational and Psychological Measurement, Autumn, 1972.

McClelland, D., Atkinson, J., Clark, R., & Lowell, E. The achievement motive. New York: Appleton Century Crofts, 1953.